

Whole genome sequencing reveals the contribution of long-term carriers in Staphylococcus aureus outbreak investigation

Article (Accepted Version)

Gordon, N C, Pichon, B, Golubchik, T, Wilson, D J, Paul, J, Blanc, D S, Cole, K, Collins, J, Cortes, N, Cubbon, M, Gould, F K, Jenks, P J, Llewelyn, M, Nash, J Q, Orendi, J M et al. (2017) Whole genome sequencing reveals the contribution of long-term carriers in Staphylococcus aureus outbreak investigation. *Journal of Clinical Microbiology*, 55 (7). pp. 2188-2197. ISSN 0095-1137

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/68044/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

1 Whole Genome Sequencing reveals the contribution of
2 long-term carriers in *Staphylococcus aureus* outbreak
3 investigation
4

5 Gordon NC^{1,2*}, Pichon B³, Golubchik T^{1,2}, Wilson DJ^{1,2}, Paul J⁴, Blanc DS⁵, Cole K⁴,
6 Collins J⁶, Cortes N⁷, Cubbon M⁸, Gould FK⁶, Jenks PJ⁹, Llewelyn M^{8,10}, Nash JQ¹¹,
7 Orendi JM¹², Paranthaman K¹³, Price J⁴, Senn L⁵, Thomas HL¹³, Wyllie S⁷, Crook
8 DW^{1,2,13,14}, Peto TEA^{1,2,14}, Walker AS^{1,2,14#}, Kearns AM^{3#}
9

10 1. National Institute for Health Research Oxford Biomedical Research Centre, John Radcliffe
11 Hospital, Oxford, UK

12 2. Nuffield Department of Medicine, University of Oxford, UK

13 3. Antimicrobial Resistance and Healthcare Associated Infections Reference Unit, Public
14 Health England, Colindale, UK

15 4. Public Health England, Royal Sussex County Hospital, Brighton, UK.

16 5. Lausanne University Hospital, Service of Preventative Medicine, Lausanne, Switzerland

17 6. Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle, UK

18 7. Portsmouth Hospitals NHS Trust, Portsmouth, UK

19 8. Brighton and Sussex University Hospitals NHS Trust, Brighton, UK

20 9. Plymouth Hospitals NHS Trust, Plymouth, UK

21 10. Brighton and Sussex Medical School, Falmer, UK

22 11. East Kent Hospitals NHS Foundation Trust, Canterbury, UK

23 12. Royal Stoke University Hospital, University Hospitals of North Midlands NHS Trust,
24 Stoke-on-Trent, UK

25 13. Public Health England, London, UK

26 14. The National Institute for Health Research Health Protection Research Unit in Healthcare
27 Associated Infections and Antimicrobial Resistance at University of Oxford
28

29 *Corresponding author; # contributed equally
30

31 *Dr NC Gordon

32 Nuffield Department of Medicine

33 John Radcliffe Hospital

34 Oxford OX3 9DU

35 Tel: +44(0)1865-220883

Fax: +44(0)1865-22195

36 E-mail: claire.gordon@ndm.ox.ac.uk

37 **Key words**

38 *Staphylococcus aureus*, whole genome sequencing, outbreaks

39

40 **Running title**

41 Whole genome sequencing for *S. aureus* outbreaks

42

43 **Abstract**

44 Whole genome sequencing (WGS) makes it possible to determine the relatedness of
45 bacterial isolates at high resolution, helping to characterise outbreaks. However, for
46 *Staphylococcus aureus*, accumulation of within-host diversity during carriage might limit
47 interpretation of sequencing data.

48 In this study, we hypothesised the converse: that within-host diversity can in fact be
49 exploited to reveal the involvement of long-term carriers (LTCs) in outbreaks. We analysed
50 WGS data from 20 historical outbreaks, and applied phylogenetic methods to assess genetic
51 relatedness and estimate time to most recent common ancestor (TMRCA). Findings were
52 compared with the routine investigation results and epidemiological evidence.

53 Outbreaks with epidemiological evidence for an LTC source had a mean estimated TMRCA
54 (adjusted for outbreak duration) of 243 days (95% CI 143-343), compared with 55 days (28-
55 81) for outbreaks lacking epidemiological evidence for an LTC ($p=0.004$). A threshold of 156
56 days predicted LTC involvement with a sensitivity of 0.875 and a specificity of 1.

57 We also found 6/20 outbreaks included isolates with differing antimicrobial susceptibility
58 profiles, however, these had only modestly increased pairwise diversity (mean 17.5 single
59 nucleotide variants (SNVs) (95% CI 17.3-17.8) vs 12.7 SNVs (12.5-12.8)) compared with
60 isolates with identical antibiograms ($p<0.0001$). Additionally, for 2 outbreaks, WGS identified
61 1 or more isolates which were genetically distinct despite having the outbreak PFGE
62 pulsotype.

63 Duration-adjusted TMRCA allowed the involvement of LTCs in outbreaks to be identified and
64 could be used to decide whether screening for long-term carriage (e.g. in healthcare
65 workers) is warranted. Requiring identical antibiograms to trigger investigation could miss
66 important contributors to outbreaks.

67

68

Introduction

To manage *Staphylococcus aureus* outbreaks effectively, infection control practitioners need to determine the relatedness of isolates from suspected cases. Whole genome sequencing (WGS) has shown superior resolution compared with standard typing techniques (*spa*, pulsed field gel electrophoresis (PFGE)) when used for individual outbreaks (1-4), and can also provide additional information about resistance, pathogenicity and population structure (5-8). However, it has been argued that the accumulation of within-host diversity during *S. aureus* carriage could result in erroneous inferences about transmission. This has been cited as a potential weakness in applying sequencing to *S. aureus* outbreaks, and may lead to misinterpretation of genuine transmission routes (1, 9, 10).

However, rather than within-host diversity being a limitation on sequencing-based outbreak investigation, it could in fact be exploited to determine whether a long-term carrier is implicated in maintaining an outbreak. This information could be used by infection control practitioners when considering whether or not to deploy extended screening (e.g. of healthcare workers).

In this study, we tested the hypothesis that WGS can be used to predict the presence of a long-term carrier as an outbreak source. First, we examined individuals with newly acquired *S. aureus* nasal carriage to ascertain whether diversity is present at acquisition or develops over time. Next, we analysed 20 *S. aureus* outbreaks, previously investigated using standard typing techniques, to assess the added utility of WGS. Finally, we compared WGS with epidemiological data to determine whether the presence of a long-term carrier maintaining the outbreak could be inferred from the WGS data.

Results

Comparison of within-host diversity in newly-acquired and long-term carriage

Eight subjects were identified with ≥ 3 consecutive bi-monthly negative nasal swabs, followed by ≥ 1 year of swabs consistently positive for *S. aureus*. All isolates were MSSA,

96 representing 7 *spa*-types, 5 sequence-types and 4 clonal complexes. Median time from first
97 to last positive sample was 490 days (range 358-727). In total, 135 isolates were
98 successfully sequenced from 16 samples. One isolate (case 1219, early sample) failed
99 quality checks and was excluded.

100

101 In 6/8 subjects, there was a significant increase in mean pairwise diversity (MPWD) between
102 the first and last samples ($p < 0.05$; figure 1). In one participant (1236) the increase was not
103 significant ($p = 0.52$), and for one (1375), there was a decrease which was marginally
104 significant ($p = 0.07$). Overall, MPWD increased from 0.88 single nucleotide variants (SNVs)
105 (95%CI 0.65-1.11) to 3.30 (2.92-3.68) between first and last samples ($p < 0.001$). Analysis of
106 the phylogenetic trees (see supplementary data) showed highly clonal early populations, and
107 in 2 participants only a single strain was observed. One individual (1219) had a more diverse
108 early sample (MPWD 4.57, 95%CI 3.10-6.04) compared with the other participants. This
109 subject's first positive swab was at month 12, and they had completed a course of co-
110 amoxiclav one day before their final negative swab. It is therefore possible that this was a
111 false negative due to antibiotic suppression, meaning that there may have been up to four
112 months of carriage prior to the first positive swab, accounting for the increased diversity.

113

114 Two participants (1218 and 1219) shared the same address, and had isolates of the same
115 *spa*-type. Participant 1219 (donor) became positive two months before participant 1218
116 (recipient). On direct comparison of both early populations, we found that the recipient had
117 an entirely clonal initial population, identical to 4/8 of the donor's strains (supplementary
118 data).

119

120 For an additional 13 participants positive at study entry, within-host diversity as measured by
121 MPWD ranged from 0 SNVs (3 individuals) to 26 SNVs. This may be due to differences in
122 acquisition time to time of first sample, which is unknown for these individuals.

123

124 **Outbreak characteristics**

125 Twenty outbreaks were included in the study (table 1). Fourteen (70%) were hospital-
126 associated: 5 neonatal units, 4 general wards, 1 surgical unit, 2 maternity units, and 2
127 involved multiple wards or hospital sites. Six (30%) were community-associated: 4
128 households, 1 nursing home and 1 school. Reasons for instigating an outbreak investigation
129 were: increase in MRSA carriage (8 outbreaks); Panton-Valentine Leukocidin (PVL)-
130 producing skin/soft tissue infection (7 outbreaks); surgical site infections (3 outbreaks);
131 MRSA bacteraemia (1 outbreak) and staphylococcal scalded skin syndrome (1 outbreak).
132 Three (15%) were due to MSSA, and 17 (85%) to MRSA. The median number of outbreak
133 cases was 7 (IQR 5-9). Median duration was 72 days (IQR 44-188).

134

135 Overall, isolates from 391 cases were sequenced. Nine (2.3%) were from health care
136 workers (HCWs), the remainder being from patients or household members. Outbreak
137 samples represented 9 clonal complexes, 11 sequence-types and 12 *spa*-types.

138

139 **Phylogenetic analysis of outbreaks**

140 Phylogenetic trees for each outbreak are provided in the supplementary data. Two outbreaks
141 had isolates which were equally or more distant than comparator isolates, despite having the
142 outbreak pulsotype: outbreak D (one isolate 53 SNVs from index case compared with 21)
143 and outbreak S (two isolates 49 and 46 SNVs from index case compared with 46). These
144 were therefore considered to be sporadic, non-outbreak isolates, and were excluded from
145 further analysis.

146

147 The overall MPWD across all outbreak sample pairs for the remaining 388 isolates was 13.8
148 SNVs (95%CI 13.6-13.9), compared with 4444 SNVs for non-outbreak *spa*-matched pairs
149 (95%CI 2492-6395) and 30192 SNVs for non-outbreak isolates from the same units (95%CI
150 29781-30603). All outbreak isolates were ≤ 30 SNVs from the index case. 381/388 (98%)
151 were ≤ 10 SNVs from their nearest neighbour. The 7 more distant isolates came from

152 outbreaks lasting more than 6 months (B, G and S). All isolates were mapped to a standard
153 reference genome: mapping to an alternative reference strain (performed for 6 outbreaks)
154 yielded only 2 additional SNVs overall (see supplementary data), with no effect on topology.

155

156 **Time to most recent common ancestor (TMRCA) and long-term carriers**

157 Twelve outbreaks (60%) had epidemiological evidence of a long-term carrier (LTC): 3
158 included cases with recurrent staphylococcal disease, in 5 an LTC was suspected due to
159 non-overlapping ward stays, and in 4, at least one case had post-outbreak long-term
160 carriage (figure 2). The pairwise distances between isolates from outbreaks with evidence
161 for an LTC ranged from 0 to 46 SNVs, compared with 0 to 10 SNVs for outbreaks with no
162 evidence for an LTC (table 2). Mean duration-adjusted TMRCA for outbreaks with a
163 suspected or proven LTC was 243 days (95% CI 143-343) compared with 55 days (28-81)
164 for outbreaks with no evidence for an LTC ($p=0.004$, figure 2). Excluding post-outbreak
165 carriage, analysis of the receiver-operating-characteristic curve gave an AUC of 0.953 (95%
166 CI 0.851-1). Using the Youden index to select the optimal threshold gave a cut-off value of
167 156 days, with a sensitivity of 0.875 and a specificity of 1.

168

169 **Relationship between PFGE pulsotype / antibiogram and SNV distance**

170 Five outbreaks contained isolates differing by ≥ 1 band from the index case on PFGE. MPWD
171 between outbreak isolates with identical PFGE pulsotypes was 13.6 SNVs (13.4-13.7),
172 compared with 17.3 (17.0-17.6) between isolates with differing pulsotypes ($p<0.0001$).

173

174 In 6/20 outbreaks, antimicrobial susceptibility differed across isolates, confirmed by the
175 presence / absence of mobile resistance determinants identified using BLAST, however,
176 these clearly belonged to the outbreak on phylogenetic analysis. MPWD between isolates
177 sharing an antibiogram was 12.7 SNVs (95% CI 12.5-12.8) compared with 17.5 (17.3-17.8)
178 for isolates with differing antibiograms ($p<0.0001$), although a substantial number of isolate
179 pairs with different antibiograms had 0 SNVs between their core genomes (figure 4).

180

181 For other factors potentially related to outbreak diversity, there was no evidence of
182 association between MPWD and any of: outbreak duration, reason for investigation,
183 epidemiological setting or MRSA phenotype ($p>0.05$).

184

185 Discussion

186 We have tested the use of WGS for *S. aureus* outbreak investigation using 20 outbreaks. By
187 comparing observed outbreak SNV distances with non-outbreak and spa/MLST specific
188 diversity, we were able to distinguish outbreak from non-outbreak strains.

189

190 Our observation of minimal diversity in recent acquisition of nasal carriage is reassuring for
191 the application of WGS data to outbreaks. For the donor-recipient pair, we observed a
192 narrow transmission bottleneck, with a clonal founding population despite a diverse donor
193 population. Although this is a single case, the findings are supported by the minimal diversity
194 seen in the early samples for the majority of carriage study subjects, and further evidence for
195 a narrow transmission bottleneck is provided by the relatively short SNV distances observed
196 across the outbreaks. Taken together, these findings suggest that, in an acute short term
197 outbreak, there will be insufficient time for diversity to accumulate.

198

199 If WGS is to be used routinely for outbreak investigation, these findings provide evidence
200 that single colony sequencing is likely to identify clusters reliably in this context, allowing
201 ease of interpretation and ensuring that WGS remains an affordable alternative to standard
202 typing, as a requirement for sequencing multiple colonies per case, as implied by previous
203 investigators (1, 10), would rapidly escalate costs and render WGS too expensive for routine
204 use.

205

206 Previous carriage studies have found greater distances than seen here (9, 11), however,
207 these did not account for estimated time of acquisition. We postulate that the existence of a

208 significant cloud of diversity (4, 12) may be a marker of long-term carriage, and therefore, in
209 outbreaks, higher diversity may indicate the involvement of an LTC, with outbreak diversity
210 reflecting the donor cloud.

211

212 In support of this, we observed a significant difference in duration-adjusted TMRCAs between
213 outbreaks with and without evidence of an LTC. The longest TMRCAs were in hospital
214 outbreaks with indirect links between cases (i.e. non-overlapping ward stays). The likelihood
215 of “missed” cases in these outbreaks was considered low due to enhanced screening, and
216 the most likely reason for the reoccurrence of the outbreak strain was thought by the
217 investigating teams to be either re-introduction from the community (outbreak G) or a staff
218 member with long-term carriage (outbreaks A, I, N and S). Staff carriage was proven in one
219 outbreak (by sampling and subsequent termination of the outbreak on their exclusion), but in
220 the remaining outbreaks HCWs were either not sampled, or HCW sampling was anonymised
221 and positive results could not be linked definitively with the suspected carrier.

222

223 The study necessarily reflects the circulating *S. aureus* clones in the UK and the concerns of
224 local infection control teams. The sampling frame is therefore enriched for MRSA and PVL-
225 positive outbreaks and neonatal unit; despite this there is a wide representation of sequence
226 types.

227

228 Despite the enhanced surveillance during each outbreak, there are inevitably missing
229 transmission links, due to missed sampling, suppression from antimicrobial therapy, or
230 delays in identifying contacts. One reason for missed samples may be the use of
231 antibiograms as an initial screening tool for identifying putative outbreak isolates, as most
232 investigating teams only collected isolates with identical or highly similar antimicrobial
233 susceptibility profiles. However, in the six outbreaks where isolates were included with
234 differing antibiograms, the core genomes were remarkably conserved. This is presumably

235 due to the ready loss/gain of mobile genetic elements (13), and shows that reliance on
236 antibiograms may lead to samples being wrongly excluded.

237

238 The variability of mobile elements is also important for interpreting genetic distances.
239 Recombination events such as gain/loss of a mobile element will introduce a large number of
240 SNVs even though this represents a single genetic event. Current analysis tools which can
241 accommodate this are computationally complex and, for large datasets, require sizable
242 computing resources. A simpler approach is to exclude the “mobile-ome” from phylogenetic
243 analyses and compare only the core genome, and the results above demonstrate that this is
244 an acceptable strategy. Similarly, mapping to alternative reference strains (performed for six
245 outbreaks) had minimal effect on SNV analysis and phylogeny, removing the need for
246 identification of clonal complex or index case assembly prior to phylogenetic analysis. This
247 streamlined approach brings WGS closer to routine use, as a readily deployable method with
248 a minimal burden of computational time and bioinformatic expertise.

249

250 In conclusion, we have demonstrated how a WGS-based approach can be applied to *S.*
251 *aureus* outbreak investigations. We have shown that current sampling strategies provide
252 sufficient information to determine whether isolates belong to an outbreak, and that, rather
253 than confounding the investigation, within-host diversity can be utilised to identify the
254 possible involvement of a long-term carrier, potentially enhancing the infection control
255 response. Combining this with directed multi-sampling of suspected LTCs (1) may be a cost-
256 effective method of using WGS to ensure that, where HCWs are implicated, potentially
257 career altering decisions are made using the best possible evidence.

258

259 **Methods**

260 **Comparison of within-host diversity in newly acquired and long-term carriage**

261 Eight participants were identified from a population study of *S. aureus* nasal carriage in
262 adults attending general practices in Oxfordshire (14), in which participants had nasal swabs

263 taken at two-monthly intervals, with positive samples stored as mixed glycerol stocks taken
264 by sweeping across multiple colonies on the primary plates to preserve the diversity of
265 carried strains (11). The eight participants were negative for nasal carriage at recruitment
266 and had consistently negative swabs for ≥ 6 months subsequently, before acquiring a strain
267 which they carried for at least one year. The first and last positive samples for each
268 individual were retrieved from the mixed glycerol stocks. Samples were plated on Columbia
269 blood agar (CBA) and incubated overnight at 37°C. For each time-point, 8 individual colonies
270 (12 for one individual, id=1218) were selected and sub-cultured to a further CBA plate and
271 again incubated overnight at 37°C.

272

273 We also retrieved sequencing data from 13 participants previously investigated, for whom
274 the approximate time of acquisition was unknown (9). Each of these had 8-12 individual
275 colonies sequenced.

276

277 **Collection of outbreak isolates and epidemiological data**

278 19 outbreaks were purposively sampled in collaboration with the Public Health England
279 (PHE) reference laboratory, representing a range of sequence-types and epidemiological
280 settings, and including both MRSA and MSSA. One further outbreak was investigated in
281 conjunction with Lausanne University Hospital, Switzerland (15, 16). Epidemiological
282 information was obtained from each infection control team (specimen date, site, ward
283 location and, where applicable, admission/discharge dates and previous screening results).

284

285 For each outbreak, additional background isolates were also included for comparison. We
286 sequenced all isolates submitted to PHE as part of the outbreak investigation, including
287 those identified as “non-outbreak” by routine typing, to estimate expected genetic diversity
288 outwith the outbreak strain, and to ensure that the apparent outbreak strains were not part of
289 an ongoing clonal expansion. We also included non-epidemiologically linked isolates

290 matched for *spa*-type and/or MLST, to provide a comparison for expected within-*spa*

291 distances, and to provide an outgroup for phylogenetic analysis.

292

293 Isolates were retrieved from single colony frozen stocks held at the PHE reference

294 laboratory, Colindale, London, or at Lausanne University Hospital. We used only the first

295 isolate from each case, and included isolates both from clinical samples and screening

296 swabs.

297

298 **Extraction and sequencing**

299 DNA was extracted and sequenced as previously described (6) from a single colony sub-

300 cultured on CBA and incubated for 18-24hrs. Sequencing was performed using the Illumina

301 HiSeq or MiSeq platforms.

302

303 **Genome assembly and construction of phylogenetic trees**

304 For all outbreaks, reads were aligned using Stampy v1.0.17 to a standard reference genome

305 (MRSA252: GenBank NC_002952) (17). Six outbreaks were also mapped to clonal-complex

306 specific reference genomes obtained from in-house collections or GenBank. Single

307 nucleotide variants were identified across all mapped non-repetitive sites using SAMtools v

308 0.1.18 mpileup, with the extended base-alignment quality flag and masking of mobile genetic

309 elements. A consensus of $\geq 75\%$, and ≥ 5 reads, including one in each direction, was required

310 to support a SNV, and calls were required to be homozygous under a diploid model.

311 Maximum likelihood trees were estimated from the mapped whole genomes using PhyML

312 (18).

313

314 **Outbreak analysis and calculation of TMRCA**

315 The index case was defined as the earliest microbiologically confirmed case in each cluster.

316 Outbreak cases were defined as those sharing related PFGE pulsotypes (19) plus a definite

317 epidemiological link to the index or secondary cases (>24 hr stay in same ward, or

household/classroom/similar community situation with prolonged contact e.g. childcare). For each outbreak case, the genetic distance in SNVs was calculated from the index case and the next nearest neighbour. If an isolate was more distant from the index case than the nearest *spa*/MLST-matched comparator, it was considered sporadic and excluded from further outbreak analysis.

323

We classified each outbreak according to the possibility of long-term carrier involvement (LTC, carrying for ≥ 6 weeks) as follows:

326

- 1) LTC not suspected: direct contact between cases, no history of pre-existing staphylococcal disease
- 2) evidence for a pre- or peri-outbreak LTC: either ≥ 1 case with prior history of recurrent staphylococcal disease, or non-overlapping hospital stays (ward case identified after a case-free interval, indicating a possible healthcare-worker carrier)
- 3) evidence of a post-outbreak LTC: ≥ 1 case with positive nasal swab > 6 weeks after initial swab (indicating a propensity for long term carriage).

334

To evaluate the relationship between outbreak diversity and the likelihood of a long term carrier, we estimated time to most recent common ancestor (TMRCA) using BEAST v1.8.1 (20). We applied a simple HKY substitution model with constant population size and a standardized substitution rate of 3.3×10^{-6} substitutions per genome per year (7) (see supplementary data). To control for differences in outbreak duration, outbreaks were censored at six months, and the (censored) outbreak duration subtracted from the calculated TMRCA to obtain a duration-adjusted TMRCA.

342

We compared SNV distances between isolates of identical pulsotype and those differing by one or more band. To determine whether there was an increase in genetic diversity

344

345 associated with acquisition of antimicrobial resistance, we also interrogated the predicted
346 antibiograms as previously described (21).

347

348 Statistical analyses were performed using Stata v13.1. Mean pairwise differences were
349 modelled using normal linear regression using robust standard errors to account for
350 dependence within person/outbreak. The ability of TMRCA to differentiate between
351 outbreaks with evidence for an LTC compared with outbreaks with no evidence for an LTC
352 was evaluated using a receiver-operating-characteristic curve analysis.

353

354 The sequences reported in this paper have been deposited in the NCBI Sequence Read
355 Archive under bioproject number PRJNA380544.

356

357 **Acknowledgements**

358 We are grateful to the staff of the PHE Staphylococcal reference service (especially Mrs
359 Marjorie Ganner) for help in identifying and locating samples. We would also like to extend
360 our thanks to the Infection Control teams at each site for permission to use isolates and
361 epidemiological data, in particular the team at East Kent Hospitals NHS Foundation Trust.

362

363 **Funding**

364 This study is supported by the Health Innovation Challenge Fund (a parallel funding
365 partnership between the Wellcome Trust [WT098615/Z/12/Z] and the Department of Health
366 [grant HICF-T5-358]), and by the National Institute for Health Research Health Protection
367 Research Unit (NIHR HPRU) in Healthcare Associated Infections and Antimicrobial
368 Resistance at University of Oxford at in partnership with Public Health England (PHE) [grant
369 number HPRU-2012-10041]. NCG is an MRC doctoral research fellow. The investigation of
370 the Swiss outbreak was supported by a grant from the Swiss Foundation for Research (no.
371 31003A_150029). DWC and TEAP are NIHR Senior Investigators. The views expressed are

372 those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of
373 Health, Wellcome Trust, the Medical Research Council or Public Health England.

References

1. Harris SR, Cartwright EJ, Torok ME, Holden MT, Brown NM, Ogilvy-Stuart AL, Ellington MJ, Quail MA, Bentley SD, Parkhill J, Peacock SJ. 2013. Whole-genome sequencing for analysis of an outbreak of methicillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Infect Dis* 13:130-6.
2. Koser CU, Holden MT, Ellington MJ, Cartwright EJ, Brown NM, Ogilvy-Stuart AL, Hsu LY, Chewapreecha C, Croucher NJ, Harris SR, Sanders M, Enright MC, Dougan G, Bentley SD, Parkhill J, Fraser LJ, Betley JR, Schulz-Trieglaff OB, Smith GP, Peacock SJ. 2012. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N Engl J Med* 366:2267-75.
3. Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, Ip CL, Wilson DJ, Didelot X, O'Connor L, Lay R, Buck D, Kearns AM, Shaw A, Paul J, Wilcox MH, Donnelly PJ, Peto TE, Walker AS, Crook DW. 2012. A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. *BMJ Open* 2.
4. Tong SY, Holden MT, Nickerson EK, Cooper BS, Koser CU, Cori A, Jombart T, Cauchemez S, Fraser C, Wuthiekanun V, Thaipadungpanit J, Hongsuwan M, Day NP, Limmathurotsakul D, Parkhill J, Peacock SJ. 2015. Genome sequencing defines phylogeny and spread of methicillin-resistant *Staphylococcus aureus* in a high transmission setting. *Genome Res* 25:111-8.
5. Alam MT, Petit RA, 3rd, Crispell EK, Thornton TA, Conneely KN, Jiang Y, Satola SW, Read TD. 2014. Dissecting vancomycin-intermediate resistance in *Staphylococcus aureus* using genome-wide association. *Genome Biol Evol* 6:1174-85.
6. Young BC, Golubchik T, Batty EM, Fung R, Larner-Svensson H, Votintseva AA, Miller RR, Godwin H, Knox K, Everitt RG, Iqbal Z, Rimmer AJ, Cule M, Ip CL, Didelot X, Harding RM, Donnelly P, Peto TE, Crook DW, Bowden R, Wilson DJ. 2012. Evolutionary dynamics of *Staphylococcus aureus* during progression from carriage to disease. *Proc Natl Acad Sci U S A* 109:4550-5.
7. Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, Gardete S, Tavares A, Day N, Lindsay JA, Edgeworth JD, de Lencastre H, Parkhill J, Peacock SJ, Bentley SD. 2010. Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 327:469-74.
8. Holden MT, Hsu LY, Kurt K, Weinert LA, Mather AE, Harris SR, Strommenger B, Layer F, Witte W, de Lencastre H, Skov R, Westh H, Zemlickova H, Coombs G, Kearns AM, Hill RL, Edgeworth J, Gould I, Gant V, Cooke J, Edwards GF, McAdam PR, Templeton KE, McCann A, Zhou Z, Castillo-Ramirez S, Feil EJ, Hudson LO, Enright MC, Balloux F, Aanensen DM, Spratt BG, Fitzgerald JR, Parkhill J, Achtman M, Bentley SD, Nubel U. 2013. A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Res* 23:653-64.
9. Golubchik T, Batty EM, Miller RR, Farr H, Young BC, Larner-Svensson H, Fung R, Godwin H, Knox K, Votintseva A, Everitt RG, Street T, Cule M, Ip CL, Didelot X, Peto TE, Harding RM, Wilson DJ, Crook DW, Bowden R. 2013. Within-host evolution of *Staphylococcus aureus* during asymptomatic carriage. *PLoS One* 8:e61319.
10. Worby CJ, Lipsitch M, Hanage WP. 2014. Within-host bacterial diversity hinders accurate reconstruction of transmission networks from genomic distance data. *PLoS Comput Biol* 10:e1003549.
11. Votintseva AA, Miller RR, Fung R, Knox K, Godwin H, Peto TE, Crook DW, Bowden R, Walker AS. 2014. Multiple-strain colonization in nasal carriers of *Staphylococcus aureus*. *J Clin Microbiol* 52:1192-200.

- 423 12. Paterson GK, Harrison EM, Murray GG, Welch JJ, Warland JH, Holden MT, Morgan FJ, Ba X,
424 Koop G, Harris SR, Maskell DJ, Peacock SJ, Herrtage ME, Parkhill J, Holmes MA. 2015.
425 Capturing the cloud of diversity reveals complexity and heterogeneity of MRSA carriage,
426 infection and transmission. *Nat Commun* 6:6560.
- 427 13. McCarthy AJ, Loeffler A, Witney AA, Gould KA, Lloyd DH, Lindsay JA. 2014. Extensive
428 horizontal gene transfer during *Staphylococcus aureus* co-colonization in vivo. *Genome Biol*
429 *Evol* 6:2697-708.
- 430 14. Miller RR, Walker AS, Godwin H, Fung R, Votintseva A, Bowden R, Mant D, Peto TE, Crook
431 DW, Knox K. 2014. Dynamics of acquisition and loss of carriage of *Staphylococcus aureus*
432 strains in the community: the effect of clonal complex. *J Infect* 68:426-39.
- 433 15. Vogel V, Falquet L, Calderon-Copete SP, Basset P, Blanc DS. 2012. Short term evolution of a
434 highly transmissible methicillin-resistant *Staphylococcus aureus* clone (ST228) in a tertiary
435 care hospital. *PLoS One* 7:e38969.
- 436 16. Senn L, Clerc O, Zanetti G, Basset P, Prod'homme G, Gordon NC, Sheppard AE, Crook DW, James
437 R, Thorpe HA, Feil EJ, Blanc DS. 2016. The Stealthy Superbug: the Role of Asymptomatic
438 Enteric Carriage in Maintaining a Long-Term Hospital Outbreak of ST228 Methicillin-
439 Resistant *Staphylococcus aureus*. *MBio* 7.
- 440 17. Holden MT, Feil EJ, Lindsay JA, Peacock SJ, Day NP, Enright MC, Foster TJ, Moore CE, Hurst L,
441 Atkin R, Barron A, Bason N, Bentley SD, Chillingworth C, Chillingworth T, Churcher C, Clark L,
442 Corton C, Cronin A, Doggett J, Dowd L, Feltwell T, Hance Z, Harris B, Hauser H, Holroyd S,
443 Jagels K, James KD, Lennard N, Line A, Mayes R, Moule S, Mungall K, Ormond D, Quail MA,
444 Rabinowitch E, Rutherford K, Sanders M, Sharp S, Simmonds M, Stevens K, Whitehead S,
445 Barrell BG, Spratt BG, Parkhill J. 2004. Complete genomes of two clinical *Staphylococcus*
446 *aureus* strains: evidence for the rapid evolution of virulence and drug resistance. *Proc Natl*
447 *Acad Sci U S A* 101:9786-91.
- 448 18. Guindon S, Delsuc F, Dufayard JF, Gascuel O. 2009. Estimating maximum likelihood
449 phylogenies with PhyML. *Methods Mol Biol* 537:113-37.
- 450 19. Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, Persing DH, Swaminathan B.
451 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel
452 electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol* 33:2233-9.
- 453 20. Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees.
454 *BMC Evol Biol* 7:214.
- 455 21. Gordon NC, Price JR, Cole K, Everitt R, Morgan M, Finney J, Kearns AM, Pichon B, Young B,
456 Wilson DJ, Llewelyn MJ, Paul J, Peto TE, Crook DW, Walker AS, Golubchik T. 2014. Prediction
457 of *Staphylococcus aureus* antimicrobial resistance by whole-genome sequencing. *J Clin*
458 *Microbiol* 52:1182-91.

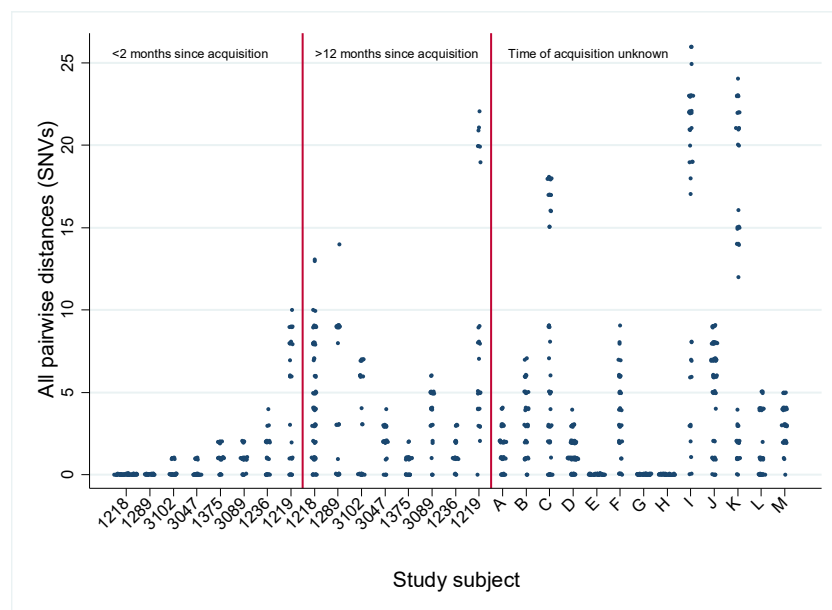
459

460 **Tables and Figures**461 **Table 1.** Description of 20 outbreaks analysed by whole genome sequencing

Outbreak	Epidemiological category	No of cases	Reason for outbreak investigation	MRSA or MSSA	Clonal complex	MLST	<i>spa</i>	Duration (days)	PFGE pulsotypes	Outbreak antibiograms
A	Hospital - general ward	5	MRSA colonisation	MRSA	CC22	ST22	t032	367	All identical	All identical
B	Hospital - general ward	6	<i>S. aureus</i> wound infections	MSSA	CC8	ST2021	t008	412	All identical	All identical
C	Hospital - general ward	7	<i>S. aureus</i> wound infections	MRSA	CC8	ST239	t037	98	All identical	All identical
D	Hospital - general ward	17	MRSA colonisation	MRSA	CC8	ST8	t008	88	All identical	All identical
E	Hospital – surgical unit	8	<i>S. aureus</i> wound infections	MRSA	CC22	ST22	t022	18	All identical	All identical
F	Hospital - multiple wards	50	MRSA colonisation	MRSA	CC5	ST228	t041	122	2 pulsotypes	All identical
G	Hospital - multiple wards	187	MRSA colonisation	MRSA	CC8	ST8	t008	454	4 pulsotypes	3 antibiograms
H	Hospital - maternity unit	6	PVL-related SSTIs	MRSA	CC1	ST772	t657	70	All identical	All identical
I	Hospital - maternity unit	9	Scalded skin syndrome	MSSA	CC15	ST2434	t346	70	All identical	2 antibiograms
J	Hospital - neonatal unit	3	MRSA colonisation	MRSA	CC59	ST59	t216	8	All identical	All identical
K	Hospital - neonatal unit	4	MRSA colonisation	MRSA	CC22	ST22	t5892	43	All identical	All identical
L	Hospital - neonatal unit	6	MRSA colonisation	MRSA	CC30	ST30	t019	57	All identical	All identical
M	Hospital - neonatal unit	8	MRSA bacteraemia	MRSA	CC88	ST88	t5973	65	All identical	3 antibiograms
N	Hospital - neonatal unit	41	MRSA colonisation	MRSA	CC22	ST22	t5892	1526	All identical	2 antibiograms
O	Household	3	PVL-related SSTIs	MRSA	CC30	ST30	t019	8	All identical	All identical
P	Household	4	PVL-related SSTIs	MRSA	CC30	ST30	t019	20	3 pulsotypes	All identical
Q	Household	5	PVL-related SSTIs	MRSA	CC30	ST30	t019	195	2 pulsotype	All identical
R	Household	8	PVL-related SSTIs	MRSA	CC30	ST30	t019	44	All identical	2 antibiograms
S	Nursing home	9	PVL-related SSTIs	MRSA	CC30	ST30	t019	298	2 pulsotypes	3 antibiograms
T	School	5	PVL-related SSTIs	MSSA	CC121	ST121	t645	74	All identical	All identical

462 PVL: Panton-Valentine Leukocidin; MLST: multi-locus sequence-type

463 **Figure 1.** All pairwise differences between early (<2 months since acquisition) and late (>12
464 months since acquisition) nasal swab samples from 7 patients with previous negative nasal
465 swabs. Included for comparison are samples from patients positive at entry to the study
466 (time of acquisition unknown).
467



468

469

470 **Table 2:** Long term carrier category, duration-adjusted TMRCA and SNV range for outbreaks
 471 investigated using WGS

Outbreak	Long term carrier category	Duration- adjusted TMRCA, days (95% highest posterior density interval)	Range of distances between all isolates in cluster, SNVs
A	Indirect ward contact	285 (134-445)	0-19
B	Indirect ward contact	515 (394-646)	0-24
C	Direct ward contact	103 (61-228)	0-9
D	Direct ward contact	78 (21-153)	0-10
E	Post outbreak LTC	96 (40-165)	0-6
F	Direct ward contact	86 (36-160)	0-5
G	Indirect ward contact	394 (259-539)	0-46
H	Direct ward contact	46 (5-95)	0-4
I	Post outbreak LTC	30 (1-69)	0-5
J	Post outbreak LTC	161 (61-271)	0-9
K	Direct ward contact	31 (0-82)	1-4
L	Post outbreak LTC	216 (134-314)	0-8
M	Direct ward contact	54 (9-107)	0-4
N	Indirect ward contact	156 (50-275)	0-36
O	Pre-outbreak LTC	378 (240-531)	9-25
P	Pre-outbreak LTC	54 (12-124)	1-11
Q	Pre-outbreak LTC	204 (108-306)	1-13
R	Household	29 (0-73)	1-2
S	Indirect ward contact	431 (269-599)	3-32
T	School	12 (0-40)	0-2

472

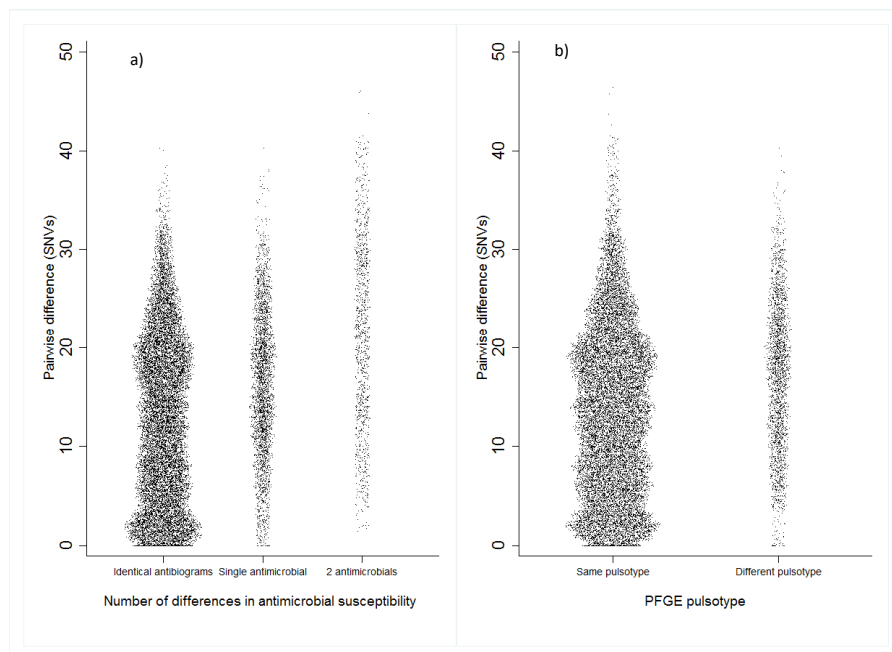
473

477



479

480 **Figure 3.** Pairwise SNV differences for all pairs within an outbreak, where isolates had differing antibiograms (a) or differing PFGE pulsotypes (b)



481

